Abstract #5422

Machine Learning Models Identify Key Histological Features of Renal Cell Carcinoma Subtypes

Samuel Vilchez*, Isaac Finberg*, Miles Markey*, Shima Nofallah*, Kathleen Sucipto, Fedaa Najdawi, Geetika Singh, Ben Trotter, Victoria Mountain, Jake Conway, Robert Egger, Chintan Parmar, Ilan Wapinski, Stephanie Hennek, Jon Glickman

*These authors contributed equally to this project

PathAI, Boston, MA

BACKGROUND

Renal Cell Carcinoma (RCC) is a heterogeneous disease that can be classified into subtypes for diagnosis based on assessment of tumor histology, and multiple molecular biomarkers and mutations .[1, 2]

Clear cell (cc) RCC is the predominant subtype comprising 80% of all cases, with papillary and chromophobe carcinoma accounting for 80% of all other RCCs, but overall, 16 different RCC subtypes have been identified.[1] Treatment selection and prognosis varies by subtype, and treatment response has been associated with the cell and tissue composition of the tumor microenvironment (TME), including tumor-specific mutations.[1] For example, *PBRM1* is commonly mutated gene in RCC that may contribute to disease prognosis and response to immunotherapy, although its role is currently unclear.[3-6] Exhaustive classification and quantification of the TME by machine learning (ML) models has the potential to reveal associations between tumor histology and mutations or molecular biomarkers.[7]

Here, ML models quantified histologic features of the TME directly from RCC hematoxylin and eosin (H&E)-stained whole slide images (WSI). The potential for model outputs to predict clinically-relevant biomarkers was investigated.

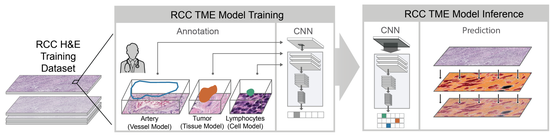
CONCLUSIONS

- ML model-quantification of RCC TME histology allowed identification of spatially specific differences that correlate with histological subtypes, vascularization, and mutations in *PBRM1*
- *PBRM1* was found to correlate with an increase in lymphocytes and a decrease in macrophages in the cancer epithelium
- Previous reports of *PBRM1* association with a low abundance of immune cells were conducted using bulk transcriptomics[3,8] where accurate assessment of the relative abundance of TME components is dependent on on tumor purity and lack of spatial information may affect interpretation
- ML models enable direct quantification of the TME allowing for granular insights into connections between histology and mutations that may play a role in disease prognosis or response to therapy
- Further support for our findings by evaluation of additional cohorts and comparison with patient outcomes is underway
- Complementary ML-based TME assessment and genomic analyses may be used, after further validation, to explore novel biomarkers

METHODS

WSI of RCC and non-RCC kidney biopsies and resections were annotated to train convolutional neural networks (CNNs) to quantify TME histological features. The training data set consisted of 3,208 WSIs from the cancer genome atlas (TCGA), proprietary, and commercial sources and over 100,000 annotations of cells, tissues, and vessels provided by expert pathologists (**Figure 1**). Training resulted in Cell Models that identified cancer, endothelial, and immune cells, as well as erythrocytes and fibroblasts, Tissue Models that measured areas of tumor, cancer epithelium, and stroma, and a Vessel Model that classified blood vessels. Combined, the models produced an exhaustive characterization of the RCC TME.

Figure 1. RCC TME HIF Model Development



CNN-based models were trained using WSI annotated by expert pathologists to indicate relevant histological features of the tumor microenvironment, including Cell, Tissue, and Blood Vessel features (e.g. lymphocytes, tumor stroma, and blood vessel arteries).

Thousands of human interpretable features (HIFs) were extracted from model predictions, providing a quantitative characterization of the TME across each WSI (*e.g.*, cell density within a tissue region; **Figure 2C**). 657 TCGA WSIs were utilized for downstream analysis of HIFs after sample post-processing to remove duplicates and outliers.

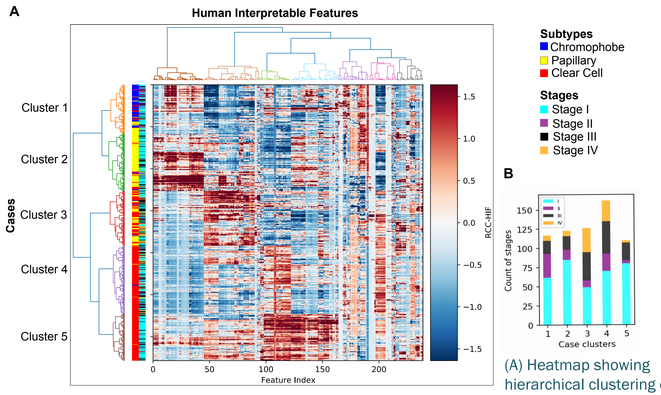
Hierarchical clustering was used to group samples based on HIF values.

Associations between HIFs and *PBRM1* loss of function (LOF) mutations and *VEGFA* mRNAseq expression in ccRCC were determined using univariate logistic regressions and Spearman's Rank correlations, respectively. False discovery rate in multiple hypothesis testing was controlled using an Empirical Brown's Method and the Benjamini-Hochberg procedure.

RESULTS

HIF hierarchical clustering ($p<10^{-6}$, chi-squared test) directly extracted chromophobe (Cluster 1), papillary (Cluster 2) and high and low-stage ccRCC (Cluster 4 and Cluster 5 respectively) and revealed a distinct cluster (Cluster 3) that was enriched for higher stage papillary and ccRCC tumors (**Figure 2 A, B, C**). Comparing the relative abundance of HIFs in each cluster, differences in the cell, tissue, and vascular composition of the TME were apparent. Subtypes could be distinguished by features describing immune cells and vascularization (**Figure 3A, C**). Cluster 3 was associated with increased prevalence of sarcomatoid regions and immune cells (**Figure 3 A, B**).

Figure 2. Hierarchical Clustering of HIFs Identifies 5 RCC Subtype Clusters



C

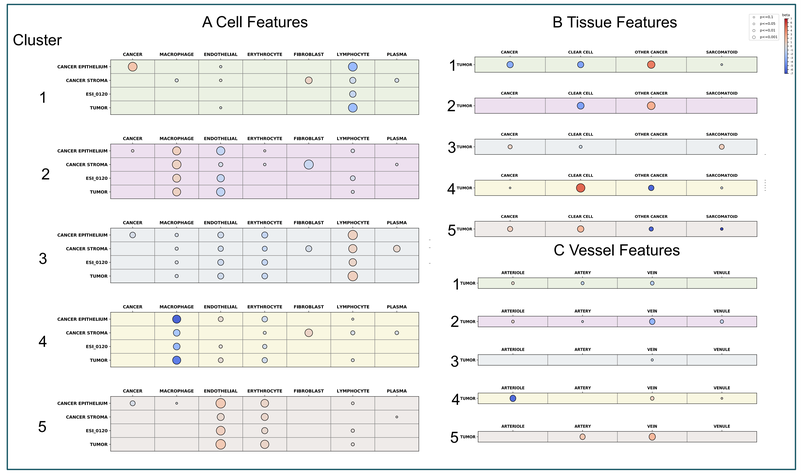
Cluster	Representative Feature
1	Number of hemosiderin laden macrophages over endothelial cells in stroma
2	Number of all predicted cells within 40µm of a lymphocyte over all predicted cells in tumor
3	Area of clear cell cancer epithelium over cancer epithelium in tissue
4	Number of erythrocyte cells over lymphocytes in cancer epithelium
5	Number of cancer cells over endothelial cells in cancer epithelium
6	Area of tumor over necrosis in tissue
7	Area of necrosis over cancer epithelium in tissue

(A) Heatmap showing hierarchical clustering of HIFs divides RCC subtypes into five clusters

(B) Distribution of cases by disease stage within cluster

(C) Representative features from each HIF cluster (color corresponds to the dendrogram color in 2A).

Figure 3. Abundance and Association of Histologic Features with RCC Subtype Clusters



Association of RCC subtype clusters with TME histological features identified by Cell (A), Tissue (B), and Vessel (C) Models

Table 1. Representative HIFs Associated with *PBRM1*

Representative HIF	Uncorrected p-value	Corrected p-value
Proportion of the number of lymphocytes within 40µm radius of an erythrocyte, over the number of immune cells within 40µm radius of an erythrocyte in cancer epithelium	6.87x10-4	0.004
Proportion of the number of endothelial cells within a 40µm radius of a hemosiderin laden macrophage, over the number of endothelial cells in cancer epithelium	0.001	0.004
Proportion of the number of endothelial cells over the number of lymphocytes in cancer epithelium	0.032	0.084
Proportion of the area of tumor over the area of necrosis in tissue	0.149	0.298
Proportion of the number of cancer cells, over the number of hemosiderin laden macrophages in cancer epithelium	0.194	0.310
Proportion of the area 120µm into the epithelial-stromal interface, over the area of cancer stroma in tissue	0.423	0.423
Proportion of the number of erythrocytes, over the number of lymphocytes in cancer stroma	0.320	0.389
Proportion of the number of all predicted cell types within a 40µm radius of a lymphocyte, over all predicted cells in cancer epithelium	0.341	0.389

Table 2. Representative HIFs Associated with *VEGFA*

Representative HIF	Uncorrected p-value	Corrected p-value	Spearman Coefficient
Proportion of the number of lymphocytes within a 40µm radius of an erythrocyte, over the number of immune cells within a 40µm radius of an erythrocyte in cancer epithelium	3.50E-12	1.22E-11	0.37
Proportion of the number of cancer cells within a 40µm radius of an endothelial cell, over cancer cells in cancer epithelium	8.96E-11	2.09E-10	0.31
Proportion of the area of tumor, over the area of necrosis in tissue	1.95E-08	2.73E-08	0.12
Proportion of the number of erythrocytes over the number of lymphocytes in cancer epithelium	0.09	0.09	0.06
Proportion of the number of cancer cells over the number of erythrocytes in cancer epithelium	1.25E-08	2.19E-08	0.01
Proportion of the number of hemosiderin laden macrophage, over endothelial cells in cancer stroma	3.36E-14	2.35E-13	-0.24
Proportion of the number of all predicted cells within a 40µm radius of a lymphocyte, over all predicted cells in cancer epithelium	0.0003	0.0004	-0.02

In ccRCC, associations between the TME and clinically-relevant mutations were examined using HIFs. Univariate analysis revealed significant associations between multiple HIFs and $PBRM1^{MUT}$, which was previously associated with response to immunotherapy. Those HIFs included and a spatially specific increase in lymphocytes in the cancer epithelium (FDR-corrected p=0.004) and a decrease in macrophages (FDR-corrected p=0.004) in the entire tumor area (**Table 1**). Spearman correlation showed that *VEGFA* expression, predictive of angiogenesis, was positively associated with an increased abundance of lymphocytes near erythrocytes (r=0.37; **Table 2**).

CONTACT EMAIL:

Stephanie Hennek:
stephanie.hennek@pathai.com

REFERENCES

1.Escudier et al. Ann. Oncol. 2019;30(5):706-720
2.Cimadamore et al. Transl. Androl. Urol. 2021 Mar;10(3):1506-1520
3.Liu et al. Nat Commun. 2020;11(1):2135.
4.Braun et al. JAMA Oncol. 2019;5(11):1631-1633.
5. Dizman et al. J Immunother Cancer. 2020;8(2):e000953
6. Carneiro et al. Kidney Cancer 2021 5(2): 79-92
7.Diao et al. Nat Commun. 2021;12(1):1613
8. Miao et al.. Sci Rep. 2022;12(1):20734.

ACKNOWLEDGMENTS

We would like to thank the members of PathAI Pathology Network who contributed towards this work.

SciStories LLC. https://scistories.com/ developed this poster template